Journal of Network and Computer Applications 46 (2014) 336–351

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Cross-layer QoE-driven admission control and resource allocation for adaptive multimedia services in LTE

K. Ivesic*, L. Skorin-Kapov, M. Matijasevic

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, Croatia

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form
18 July 2014
Accepted 18 September 2014
Available online 30 September 2014

Keywords:
Admission control
Resource allocation
Multimedia services
Simulation
LTE

ABSTRACT

This paper proposes novel resource management mechanisms for multimedia services in 3GPP Long Term Evolution (LTE) networks aimed at enhancing session establishment success and network resources management, while maintaining acceptable end-user quality of experience (QoE) levels. We focus on two aspects, namely admission control mechanisms and resource allocation. Our cross-layer approach relies on application-level user- and service-related knowledge exchanged at session initiation time, whereby different feasible service configurations corresponding to different quality levels and resource requirements can be negotiated and passed on to network-level resource management mechanisms. We propose an admission control algorithm which admits sessions by considering multiple feasible configurations of a given service, and compare it with a baseline algorithm that considers only single service configurations, which is further related to other state-of-the-art algorithms. Our results show that admission probability can be increased in light of admitting less resource-demanding configurations in cases where resource restrictions prevent admission of a session at the highest quality level. Additionally, in case of reduced resource availability, we consider resource reallocation mechanisms based on controlled session quality degradation while maintaining user QoE above the acceptable threshold. Simulation results have shown that given a wireless access network with limited resources, our approach leads to increased session establishment success (i.e., fewer sessions are blocked) while maintaining acceptable user-perceived quality levels.

1. Introduction

With the advent of high performance mobile technologies, the capabilities of data transmission on the move have been improved to a great extent. Together with the development of advanced mobile devices, networks like 3GPP Long Term Evolution (LTE) enable the provision of a wide range of multimedia services, like high quality video and online gaming, to mobile users. Ericsson Mobility Report has shown that the volume of mobile data traffic has almost doubled from the second quarter of 2012 to the second quarter of 2013, and further grown 10% between the second and the third quarter of 2013 (Ericsson, 2013). Further increases in mobile traffic volume are expected and capacity shortage is inevitable. This calls for new quality of experience (QoE) driven resource management mechanisms, as the demand and the popularity of advanced multimedia services aiming to meet user experience requirements are expected to grow (Agboma and Liotta, 2012; Baraković and Skorin-Kapov, 2013).

Many parameters affect the end user perceived multimedia service quality. At the application layer, examples of such parameters include media type, display resolution, codec type, and frame rate. At the network layer, typical parameters to consider are delay, jitter, loss, and throughput. A joint consideration and mapping of these parameters to achievable QoE levels is necessary when making intelligent resource allocation decisions. Resource management mechanisms often focus on lower layer data and Quality of Service (QoS) mechanisms, while neglecting the impact of application layer parameters. For example, typical resource allocation mechanisms may perform high-level traffic classification and prioritization, without considering the knowledge of how given mechanisms actually impact end user QoE, e.g., Nasser and Guizani (2010) classify all services into two categories and distinguish them only by their priorities. Utility functions have been commonly used in the context of formalizing the correlation between network performance and user perceived quality, expressing users degree of satisfaction with respect to corresponding multi-criteria service performance (Reichl et al., 2011).

In our previous work, we have proposed a QoS negotiation and adaptation framework for multimedia sessions (Skorin-Kapov et al., 2007; Skorin-Kapov and Matijasevic, 2009), which has

* Corresponding author.
E-mail address: krunoslav.ivesic@fer.hr (K. Ivesic).

http://dx.doi.org/10.1016/j.jnca.2014.09.010
1084-8045/

supported the application-level signalling and end-to-end negotiation of a so-called *Media Degradation Path* (MDP) to specify a mapping between session parameters, resource requirements, and corresponding user experience quality levels. The MDP has been defined as an ordered collection of different feasible session *configurations*, where each configuration specifies service operating parameters (e.g., codec, video resolution), corresponding resource requirements of each involved media flow (e.g., required bandwidth), as well as an expected *utility* value. The utility value represents a numerical indicator of user perceived quality with the regarding configuration. For example, consider an audiovisual service that supports different bitrates and encoding types, offering the audio and video flows at different quality levels (e.g., which can be chosen from depending on the type of access network an end user is using, terminal capabilities, operator policy, user subscription type, etc.). In each MDP, there is an optimal configuration (resulting in the highest achievable utility value) and several alternative ones, ordered by decreasing utility value.

During calculation of the MDP, *user* and *service* profiles are taken as input, specifying parameters such as service adaptation capabilities (service profile) and user preferences, e.g., regarding media flow importance and preferences (user profile). In this way, knowledge about the user and the service is embedded in the MDP, and it specifies a "recipe" for controlled degradation of the session to be applied in the case of resource shortage. By *controlled degradation* we refer to the act of switching to an alternative pre-specified session configuration in light of reduced resource availability, rather than experiencing reduced session quality due to uncontrolled degradation of all flows, or relying only on adaptation mechanisms initiated by the application itself. While controlled client-initiated application adaptation has been previously proposed, such as in the context of HTTP Adaptive Streaming (Oyman and Singh, 2012), we focus on network-based mechanisms that consider multiple sessions and domain-wide optimized resource allocation.

In this work we present the application of the MDP to resource management mechanisms, namely (1) admission control, and (2) resource reallocation in case of reduced resource availability, such as that caused by congestion in the network. When a new multimedia session is starting, alternative configurations from the MDP can be used if currently available resources are not sufficient for the optimal configuration, thus increasing admission probability while keeping user satisfaction at an acceptable level, since the alternative configurations reflect user preferences and acceptable quality levels. Additionally, alternative configurations degrade different flows in a different manner, depending on the user preferences. We have introduced this concept in our previous work (Ivesic et al., 2013), which we now extend with more detailed simulation tests used for evaluation of the approach. Additionally, we deal with the problem of resource reallocation in case of resource shortage. Since the multimedia sessions that we consider can, in certain occasions, increase their resource consumption considerably, in comparison to the resources assigned at the admission time, we propose a resource reallocation mechanism in case of reduced resource availability, which conducts graceful degradation by switching active sessions to less resource demanding, but agreed-on configurations. We also build on earlier results (Ivesic et al., 2010, 2011) with extensive validation in a simulated LTE access network. For evaluation purposes, we utilize the simulator tool called ADAPTISE (ADmission control and resource Allocation for adaPtive mulTImedia SErvices) developed by our group at the University of Zagreb (Ivesic et al., 2010, 2011), and the LTE-Sim tool developed by Politecnico di Bari, Italy, for performance evaluation in an LTE network (Piro et al., 2011).

Our tests show that by employing an approach considering MDP calculation, admission probability is increased as compared to a baseline admission control algorithm not considering an MDP. Additionally, we show how our baseline algorithm compares with several works from the literature, thus deriving the general conclusion that the MDP-based algorithm outperforms algorithms that do not consider different service configurations. Furthermore, results show that in cases of reduced resource availability, selected sessions are adapted in a controlled manner while maintaining acceptable quality levels and freeing resources for new incoming sessions. To the best of our knowledge, current literature lacks approaches that perform degradations to alternative configurations that reflect user- and service-related knowledge.

In order to clarify the applicability of the work proposed in this paper, we briefly discuss a possible mapping of our proposed resource management approaches to the standardized LTE Evolved Packet Core (EPC) architecture. The paper is organized as follows. In Section 2 we give a survey of related work regarding admission control and resource allocation, as well as a brief overview of resource management mechanisms in LTE networks. Section 3 presents our approaches for resource management based on user- and service-related knowledge. The simulations and the analysis of results are given in Sections 4 and 5, respectively. Section 6 provides a discussion of the proposed approach in the context of standardized LTE QoS management mechanisms, and provides concluding remarks and outlook for future work. We end the paper with a summary of contributions and outlook for ongoing and future work.

2. Related work

2.1. Admission control mechanisms

The basis of many modern admission control algorithms in cellular mobile networks has been established by Hong and Rappaport (1986), who introduced the notion of the *guard channel* as a mechanism for ensuring capacity for handoff calls by reserving a certain number of channels exclusively for them. With the introduction of different user and service categories, other schemes have been introduced, often dividing the available capacity into several zones. For example, in a scheme proposed by Nasser and Guizani (2010) there are two call categories with different priorities and the capacity is divided into four zones. The first zone is available to all calls, the second one to lower category handoff calls and all higher category calls, the third to all higher category calls, and the last one to higher category handoff calls only. Another approach suggested by Shuaibu et al. (2011) separates the available capacity based on the bit rate type: there are zones for constant bit rate and zones for variable bit rate services. Additionally, a portion of capacity reserved for handoff sessions can be dynamically adjusted. In either of the cases there is no consideration of multiple media flows or alternative session configurations. Additionally, there are many admission control algorithms with specific purpose in the literature, e.g., Wang and Qiu (2013) proposed an algorithm specially designed for LTE femtocells.

An approach that enables admission of sessions with different assigned bandwidth values has been proposed by Gudkova and Samouylov (2012). Two call types are assumed, namely voice call and video call, and there are two bit rate values for the video calls: guaranteed, and maximum bit rate. In case of insufficient capacity for an incoming voice call, video calls that have been assigned a maximum bit rate can be degraded to their lower, guaranteed bit rate. Chowdhury et al. (2013) also propose degradation in case of insufficient capacity: non real-time calls can be degraded to admit more real-time ones with separate degradation thresholds for new and real-time handoff calls. The degradation, however, does not consider user preferences in either of the cases. An approach suggested by Posoldova and Oravec (2013) considers session requirements (protocol data unit dropping probability, throughput